A New Continuous Discrete Unscented Kalman Filter

Knudsen, Torben; Leth, John-Josef

Published in:
I E E E Transactions on Automatic Control

DOI (link to publication from Publisher):
10.1109/TAC.2018.2867325

Publication date:
2019

Document Version
Accepted author manuscript, peer reviewed version

A New Continuous Discrete Unscented Kalman Filter

Torben Knudsen and John Leth

***Abstract*— The time and measurement update for the discrete time Kalman filter can be formulated in terms of conditional means and covariances. The unscented Kalman filter can be interpreted as calculating these conditional means and covariances by using the unscented transform. This approach can also be directly applied to nonlinear models as an alternative to the discrete time extended Kalman filter. In this paper, a novel method for computing the unscented Kalman filter for a nonlinear model with continuous time dynamics and discrete time measurements is presented. Compared to the existing approaches, this method is far simpler and less computationally demanding, and it performs at least as well.**

***Index Terms*— Estimation and filtering, Unscented Kalman filter, Stochastic differential equation, Unscented transform, Continuous discrete estimation.**

I. Introduction

Dynamical systems can be modeled as state space models, i.e., a first-order vector ordinary differential equation (ODE) and a static output equation. If there are measurement noise or unknown inputs, they can be modeled as (white noise) stochastic processes. Such a model is called a stochastic differential equation (SDE). Normally, the output vector and the possible input are assumed to be the only signals that can be measured. Based on these measurements, a state estimate is needed in many applications, e.g., forecasting, control, fault detection and system identification.

If both measurements and dynamics are in discrete time and the system is linear, the well-known Kalman filter [1], [2] is the optimal solution in the mean square error sense. This is abbreviated DD-KF, standing for discrete-discrete Kalman filter. A corresponding optimal solution called the CD-KF [3] exists for the CD (continuous-discrete) linear problem where the dynamic state is described by an SDE, and the measurements are at discrete times. For both DD and CD problems with nonlinear systems, the extended KF (EKF) has been very useful. The EKF simply linearizes the nonlinear system and uses the linearized system parameters where necessary in the KF. For DD nonlinear problems, [4] presents the so-called unscented KF (UKF), where the need for linearization is avoided by estimating the necessary covariances using the unscented transform (UT) [5], [6] and then directly using them in the DD-KF. This method is referred to as the discrete-discrete unscented Kalman filter (DD-UKF). One of the pioneers in this field compares DD-UKF to DD-EKF as follows: "It is more accurate, easier to implement, and uses the same order of calculations as linearization." [7, p. 401]. A similar statement is found in [8, p. 158]. This is also our experience.

Inspired by the superior performance of DD-UKF, an algorithm, referred to as CD-UKF in this paper, has been suggested by [9] for computing the UKF in the continuous-discrete case. An earlier alternative to CD-UKF is the ensemble methods (EnKF) discussed by, e.g., [10] and [11] where Monte Carlo methods are used instead of UT methods, which are much more computationally demanding.

It is typical to split the state estimation into a time update and a measurement update. This approach is done in [9] and in this paper (see section II). The measurement update for CD-UKF is similar to that for DD-UKF. The difficulties lie in the time update. The method for the time update in [9] consists of the following steps:

1) Start with the CD system.
2) Obtain a DD system by discretizing the CD system using a small time step Δt.
3) Set up the DD-UKF equations for the above system.
4) Let the time step Δt in the above equations tend to zero, resulting in a set of ordinary differential equations that then become the CD-UKF equations in [9].
5) The above so-called Moment Differential Equations (MDEs) [12] cannot be solved analytically; hence, a numerical method has to be used to integrate from one measurement time to the next.

See [9] for more details, especially regarding the time update by equations (34), (30) and (25) and the measurement update (27), where these equation numbers refer to [9].

The method of [9] is further elaborated by [13], [14]. In [15], a solution based on [9] is provided for a CD system extended with algebraic states. In [16], the time update is performed using the method from [9]; however, for the measurement update, alternative factorizations are suggested and analyzed. The most detailed analysis and development of methods for solving the MDEs numerically is found in the recent work [12]. The paper develops an adaptive solver with an automatic global error control and demonstrates its good performance.

In summary, a number of methods for solving the CD-UKF estimation problem numerically have been suggested. However, all of the above are based on solving the MDEs developed by [9].

For comparison with the method of [9], the time update steps of the method presented in this paper are as follows:

1) Start with the CD system.
2) Obtain a DD system by discretizing the CD system using a small time step Δt.
3) Use the unscented transform directly for the above system to calculate the necessary covariances and variances to advance from one measurement time to the next.

This is far simpler than the method of [9].

The main contribution of this paper is the new and simple CD-UKF method, which avoids the MDEs used in [9]. It is simpler to explain in an academic context, and it can be explained to practitioners without involving complicated mathematics related to SDEs.

Moreover, we remark that if it is applied to the Lorenz system, the CD-UKF method presented in this paper produces state and output estimates with the same precision as the CD-UKF method in [9] (see Table III) but outperforms it with respect to the accuracy of variance of the state estimate error (see Table IV). Finally, for this example, it runs more than 9 times faster than the CD-UKF method in [9] (see Table III). Using the results reported in [12], this difference might be reduced.

Notation: In what follows, the superscript $\mathtt{T}$ indicates transposition, and $\triangleq$ denotes "defined by". For random vectors x, y and z, the conditional covariance matrix and conditional variance are denoted

$$\mathrm{Cov}(x, y|z) \triangleq \mathrm{E}\left((x - \mathrm{E}(x|z))(y - \mathrm{E}(y|z))^{\mathtt{T}}|z\right) ,$$
$$\mathrm{Cov}(x|z) \triangleq \mathrm{Cov}(x, x|z) ,$$

Torben Knudsen and John Leth are with section of Automation and Control, Aalborg University, Fredrik Bajers Vej 7C2-212 DK-9220 Aalborg, Denmark {tk,jjl}@es.aau.dk

where in the unconditional case, these are occasionally denoted by $C_{xy} = \mathrm{Cov}(x, y)$ and $C_x = \mathrm{Cov}(x, x)$, respectively. Finally, the estimate of a value x is denoted by $\hat{x}$, and the estimation error by $\tilde{x} \triangleq x - \hat{x}$.

The remainder of this paper starts by presenting CD-KF for linear systems in section II. Section III presents CD-EKF for nonlinear systems. The UT is presented separately in section IV. The main contribution in the form of the new CD-UKF is developed in section V and compared with the alternative CD-UKF by [9] and CD-EKF in section VI based on a simulated Lorenz attractor example. Finally, section VII gives the conclusion.

II. The continuous–discrete KF

For reference, the basic CD-KF for linear CD systems is first presented.

Consider the linear time varying CD system:

$$dx(t) = (F(t)x(t) + B(t)u(t))\,dt + dw(t)\ , \tag{1a}$$
$$y(t_k) = H(t_k)x(t_k) + D(t_k)u(t_k) + v(t_k)\ , \tag{1b}$$
$$w(t) \in W(Q(t))\ ,\ v(t_i) \in \mathrm{N}(\underline{0}, R(t_i))\ , \tag{1c}$$
$$\mathrm{E}\left(v(t_i)v(t_j)^{\mathsf{T}}\right) = \Delta(t_i - t_j)R(t_i)\ , \tag{1d}$$

where $x \in R^n$ is the state, $u \in R^m$ is the input, $w \in R^n$ is a Wiener process, $y \in R^p$ is the output, and $v \in R^p$ is the Gaussian measurement noise. The matrices F, B, H, D are state space model parameters of suitable dimensions, Q is the incremental process noise covariance, R is the measurement noise covariance, and Δ is the discrete Dirac delta function. We remark that the SDE given by (1a) is to be understood in the sense of Ito (see [17]).

Given the measurements and the initial values,

$$Y_0^k \triangleq y(t_0), y(t_1), y(t_2) \ldots, y(t_k)\ , \tag{2a}$$
$$\hat{x}(t_0^-) = \hat{x}_0\ ,\ P(t_0^-) = P_0\ , \tag{2b}$$

the estimation problem is to find the estimate $\hat{x}(t_k)$ that minimizes the mean square error (MSE)

$$\mathrm{E}\left((x(t_k) - \hat{x}(t_k))^{\mathsf{T}} A (x(t_k) - \hat{x}(t_k))\right)\ , \tag{3}$$

for any positive semi-definite A. It is well known that the general solution, which is also applicable to nonlinear systems, is the conditional mean value

$$\hat{x}(t_k) = \mathrm{E}\left(x(t_k)\middle|Y_0^k\right)\ . \tag{4}$$

The CD-KF that gives the solution to the estimation problem for the linear time-varying CD system (1)-(2) can be formulated algorithmically as (5)-(6) below [3, Theo. 7.1]:

The measurement update at time t_k is performed as follows:

$$K(t_k) = P(t_k^-)H(t_k)^{\mathsf{T}} \left(H(t_k)P(t_k^-)H(t_k)^{\mathsf{T}} + R(t_k)\right)^{-1}\ , \tag{5a}$$
$$\hat{x}(t_k^+) = \hat{x}(t_k^-) + K(t_k)\left(y(t_k) - H(t_k)\hat{x}(t_k^-) - D(t_k)u(t_k)\right)\ , \tag{5b}$$
$$P(t_k^+) = (I - K(t_k)H(t_k))P(t_k^-)(I - K(t_k)H(t_k))^{\mathsf{T}} + K(t_k)R(t_k)K(t_k)^{\mathsf{T}}\ . \tag{5c}$$

The time update from t_k to t_{k+1} is performed as follows:

$$\hat{x}(t_k) = \hat{x}(t_k^+)\ ,\ P(t_k) = P(t_k^+)\ , \tag{6a}$$
$$\dot{\hat{x}}(t) = F(t)\hat{x}(t) + B(t)u(t)\ , \tag{6b}$$
$$\dot{P}(t) = F(t)P(t) + P(t)F(t)^{\mathsf{T}} + Q(t)\ , \tag{6c}$$
$$\hat{x}(t_{k+1}^-) = \hat{x}(t_{k+1})\ ,\ P(t_{k+1}^-) = P(t_{k+1})\ , \tag{6d}$$

where (6a) are the initial conditions at time t_k for the ODE (6b)–(6c) used to obtain the time-updated results (6d) for time t_{k+1}.

It is, of course, important to mention that the terms in algorithm (5)-(6) have probabilistic interpretations:

$$\hat{x}(t_k^-) \triangleq \mathrm{E}\left(x(t_k)\middle|Y_0^{k-1}\right)\ , \tag{7a}$$
$$\hat{x}(t_k^+) \triangleq \mathrm{E}\left(x(t_k)\middle|Y_0^{k}\right)\ , \tag{7b}$$
$$P(t_k^-) \triangleq \mathrm{Cov}\left(x(t_k)\middle|Y_0^{k-1}\right) = \mathrm{E}\left((x(t_k) - \hat{x}(t_k^-))(x(t_k) - \hat{x}(t_k^-))^{\mathsf{T}}\middle|Y_0^{k-1}\right)\ , \tag{7c}$$
$$P(t_k^+) \triangleq \mathrm{Cov}\left(x(t_k)\middle|Y_0^{k}\right) = \mathrm{E}\left((x(t_k) - \hat{x}(t_k^+))(x(t_k) - \hat{x}(t_k^+))^{\mathsf{T}}\middle|Y_0^{k}\right)\ . \tag{7d}$$

III. The continuous–discrete EKF

Consider now, in place of (1a) and (1b), the nonlinear model (8) below, where the drift f and output h are nonlinear but the noise v is still additive:

$$dx(t) = f(x(t), u(t), t)dt + dw(t)\ , \tag{8a}$$
$$y(t_k) = h(x(t_k), u(t_k), t_k) + v(t_k)\ . \tag{8b}$$

The EKF is then derived from the KF by using the following heuristic principle: use the nonlinear relations when possible and the linearization otherwise. Thus, the following are changed in the nonlinear setting.

The measurement update at time t_k (5b) is replaced by

$$\hat{x}(t_k^+) = \hat{x}(t_k^-) + K(t_k)(y(t_k) - h(\hat{x}(t_k^-), u(t_k), t_k))\ . \tag{9}$$

The time update from t_k to t_{k+1} (6b) is replaced by

$$\dot{\hat{x}}(t) = f(\hat{x}(t), u(t), t)\ . \tag{10}$$

In all the other equations, the linearized parameters (11) must be used.

$$F(t) \triangleq \frac{\partial f}{\partial x}(\hat{x}(t_k), u(t_k), t)\ , \tag{11a}$$
$$H(t) \triangleq \frac{\partial h}{\partial x}(\hat{x}(t_k^-), u(t_k), t)\ . \tag{11b}$$

Unless explicitly mentioned, (5) and (6) are henceforth to be understood with the changes (9)-(11).

IV. Unscented transform

The basic new idea in the UKF is to use the UT to calculate conditional means and covariances in the measurement (5) and the time update (6). How this is done for the nonlinear CD problem is explained in section V. This section first explains the UT itself.

The problem solved by the UT is the basic probabilistic problem of calculating the second-order statistics of x, y, given the second-order statistics of x and a relation f:

$$y = f(x)\ ,\ \mu_x = \mathrm{E}(x)\ ,\ C_x = \mathrm{Cov}(x)\ . \tag{12}$$

The UT can formally be written as

$$\left[\hat{\mu}_y, \widehat{C}_{yx}, \widehat{C}_y\right] = \mathrm{UT}(f, \mu_x, C_x)\ . \tag{13}$$

From the mathematical point of view, this amounts to approximating integrals of the type

$$\int_{R^n} h(z)\phi_x(z)dz\ , \tag{14}$$

where ϕ_x is the probability density function for x, and, e.g., $h = f$ in the case of $\hat{\mu}_y$.

For a nonlinear relation f, there is no exact solution for even the second-order statistics. The EKF can be interpreted as using the linearization approach:

$$y \approx f(\mu_x) + \nabla f(\mu_x)(x - \mu_x) \ ,$$

implying that

$$\hat{\mu}_y = f(\mu_x) \ , \ \widehat{C}_{yx} = \nabla f(\mu_x) C_x \ ,$$
$$\widehat{C}_y = \nabla f(\mu_x) C_x \nabla f(\mu_x)^{\mathsf{T}} \ .$$

A Monte Carlo approach would involve generating random realizations x_i based on the statistics of x and estimating the statistics for x, y as follows:

$$\hat{\mu}_y = \frac{1}{N}\sum_{i=1}^{N} f(x_i) \ , \ \hat{\mu}_x = \frac{1}{N}\sum_{i=1}^{N} x_i \ ,$$
$$\widehat{C}_{yx} = \frac{1}{N}\sum_{i=1}^{N} (f(x_i) - \hat{\mu}_y)(x_i - \hat{\mu}_x)^{\mathsf{T}} \ ,$$
$$\widehat{C}_y = \frac{1}{N}\sum_{i=1}^{N} (f(x_i) - \hat{\mu}_y)(f(x_i) - \hat{\mu}_y)^{\mathsf{T}} \ ,$$

with N chosen to be sufficiently large.

The UT can be seen as a Monte Carlo method in which the values of x_i are not randomly drawn but constructed to obtain certain features [18, p. 549-550]. These constructed x_is are called sigma points. In particular, if x is Gaussian, the eigenvectors of C_x can be used to construct the sigma points; see [5] for details. An introduction to UT can be found in, e.g., [4]. There are various versions of UT with different features. The UT used here is given below, where n is the dimension of x:

$$\lambda = 2 \ , \tag{15a}$$
$$k = \sqrt{n + \lambda} \ , \tag{15b}$$
$$u_i \triangleq \text{eigenvector } i \text{ for } C_x \ , \tag{15c}$$
$$l_i \triangleq \text{eigenvalue } i \text{ for } C_x \ , \tag{15d}$$
$$x_i = \begin{cases} \mu_x & , \ i = 0 \\ \mu_x + k u_i \sqrt{l_i} & , \ 1 \le i \le n \\ \mu_x - k u_i \sqrt{l_i} & , \ n+1 \le i \le 2n \ , \end{cases} \tag{15e}$$
$$w_i = \begin{cases} \frac{\lambda}{n+\lambda} & , \ i = 0 \\ \frac{1}{2(n+\lambda)} & , \ 1 \le i \le 2n \ , \end{cases} \tag{15f}$$
$$\hat{\mu}_y = \sum_{i=0}^{2n} w_i f(x_i) \ , \tag{15g}$$
$$\widehat{C}_{yx} = \sum_{i=0}^{2n} w_i (f(x_i) - \hat{\mu}_y)(x_i - \hat{\mu}_x)^{\mathsf{T}} \ , \tag{15h}$$
$$\widehat{C}_y = \sum_{i=0}^{2n} w_i (f(x_i) - \hat{\mu}_y)(f(x_i) - \hat{\mu}_y)^{\mathsf{T}} \ . \tag{15i}$$

Note that $\sum_{i=0}^{2n} w_i = 1$. This construction can be shown to give exact results for linear and affine functions $f(x) = Ax + b$. Here, eigenvectors are used to construct the sigma points. There are other ways, e.g., Cholesky factorization, which gives correct results for linear functions but different results for nonlinear functions, as shown in the test below.

The above UT has been tested and compared to the UT version by Rudolph van der Merwe and S. J. Julier , that uses Cholesky factorization [7], [19]. For reference, a Monte Carlo (MC) method using 10^4 random samples has also been included. The results for $y = f(x) = x^{\mathsf{T}}x$, i.e. squared length of x, are shown in Table I. The chosen input statistics are $\mathrm{E}(x) = [1; 1]$ and $\mathrm{Cov}(x) = [1 \ 1; 1 \ 2]$ (using the Matlab notation). The MC method assumes a Gaussian distribution of x. This squared length example is chosen because the theoretical results can be derived, as shown in Table I.

It is noted that the UT method (15) is correct up to third-order moments [7], [20]. That is, the estimates of $\mathrm{E}(y)$ and $\mathrm{Cov}(y, x)$ agree with the (correct) theoretical results, while the estimate of $\mathrm{Cov}(y)$ does not, as it includes 4^{th} moments. This is true only in the case considered, where f is the square length of x, which is Gaussian.

The UT method's performance is also compared to that of the MC method. The two methods perform similarly, even though the MC method uses 10^4 points compared to 5 points used by the UT methods. The slightly better performance of the Cholesky-based method compared to the UT is not the case in general, as shown in Table II.[1]

Method	$\mathrm{E}(y)$	$\mathrm{Cov}(y)$	$\mathrm{Cov}(y, x_1)$	$\mathrm{Cov}(y, x_2)$
	Estimates			
UT	5.00	39.0	4.00	6.00
UTCH	5.00	31.0	4.00	6.00
MC	5.04	33.9	3.99	6.03
Theo	5	34	4	6
	Relative error			
UT	0	0.147	0	0
UTCH	0	-0.0882	0	0
MC	0.00766	-0.00428	-0.00309	0.00467

TABLE I

ESTIMATION PERFORMANCE WITH $\boldsymbol{y = f(x) = x_1^2 + x_2^2}$, $\mathbf{E(x) = [1; 1]}$, $\mathbf{Cov(x) = [1 \ 1; 1 \ 2]}$. UT IS THE UNSCENTED TRANSFORM (15), UTCH IS THE CHOLESKY-BASED METHOD, MC IS THE MONTE CARLO METHOD, AND THEO REPRESENTS THE THEORETICAL CORRECT VALUES. THE BOTTOM PART CONTAINS ERRORS RELATIVE TO THE CORRECT RESULTS: (ESTIMATE-THEO)/THEO.

Method	$\mathrm{E}(y)$	$\mathrm{Cov}(y)$	$\mathrm{Cov}(y, x_1)$	$\mathrm{Cov}(y, x_2)$
	Estimates			
UT	36.8	6617	52.8	82.4
UTCH	32.0	3100	40.0	60.0
MC	35.3	8175	44.9	73.1
	Relative error			
UT	0.0427	-0.191	0.175	0.128
UTCH	-0.0933	-0.621	-0.110	-0.179

TABLE II

ESTIMATION PERFORMANCE WITH $\boldsymbol{y = f(x) = x_1^4 + x_2^4}$, $\mathbf{E(x) = [1; 1]}$, $\mathbf{Cov(x) = [1 \ 1; 1 \ 2]}$ AND THE MONTE CARLO (MC) METHOD USING $\mathbf{10^5}$ RANDOM SAMPLES. THE BOTTOM PART CONTAINS ERRORS RELATIVE TO THE MC METHOD: (ESTIMATE-MC)/MC.

V. CONTINUOUS–DISCRETE UNSCENTED KALMAN FILTER

The CD-UKF solves the same problem as the CD-EKF, i.e., estimating the state from the outputs of the system (8). To understand the CD-UKF, it is necessary to follow the derivation of the (linear) CD-KF to recognize that the measurement update (5b) and (5c) originates from basic relation for Gaussian variables. Assume that x, y are Gaussian vectors, then

$$\mathrm{E}(x|y) = \mathrm{E}(x) + \mathrm{Cov}(x, y)\,\mathrm{Cov}(y)^{-1}(y - \mathrm{E}(y)), \tag{16a}$$
$$\mathrm{Cov}(x|y) = \mathrm{Cov}(x) - \mathrm{Cov}(x, y)\,\mathrm{Cov}(y)^{-1}\,\mathrm{Cov}(y, x), \tag{16b}$$

[1]In Table II, the result of the MC method is used as a benchmark because calculating the necessary theoretical statistics for Table II can be very lengthy.

$$x - \mathrm{E}(x|y) \text{ and } y \text{ are independent,} \tag{16c}$$

and the MSE optimal estimate of x given y is the conditional mean $\mathrm{E}(x|y)$. Note the important feature of (16c) that the estimation error is independent of the data on which it is based. This fact, and its implications, is used extensively below, e.g., in (18).

Now, let y denote some information that is already available and z some new information that is to be obtained. The extra condition z must then also be included in (16), resulting in

$$\begin{aligned}\mathrm{E}(x|y,z) &= \mathrm{E}(x|z) \\ &\quad + \mathrm{Cov}(x,y|z)\,\mathrm{Cov}(y|z)^{-1}(y - \mathrm{E}(y|z)) \,,\end{aligned} \tag{17a}$$

$$\begin{aligned}\mathrm{Cov}(x|y,z) &= \mathrm{Cov}(x|z) \\ &\quad - \mathrm{Cov}(x,y|z)\,\mathrm{Cov}(y|z)^{-1}\,\mathrm{Cov}(y,x|z) \,.\end{aligned} \tag{17b}$$

Introducing the estimation errors

$$\tilde{y}(t_k^-) \triangleq y(t_k) - \hat{y}(t_k^-) \triangleq y(t_k) - \mathrm{E}\left(y(t_k)\middle|Y_0^{k-1}\right) \,, \tag{18a}$$

$$\tilde{x}(t_k^\pm) \triangleq x(t_k) - \hat{x}(t_k^\pm) \,, \tag{18b}$$

then yields

$$\mathrm{Cov}\left(\tilde{y}(t_k^-)\middle|Y_0^{k-1}\right) = \mathrm{Cov}(\tilde{y}(t_k^-)) \,, \tag{18c}$$

$$\mathrm{E}\left(x(t_k)\middle|\tilde{y}(t_k^-), Y_0^{k-1}\right) = \mathrm{E}\left(x(t_k)\middle|Y_0^{k}\right) = \hat{x}(t_k^+) \,, \tag{18d}$$

$$\begin{aligned}&\mathrm{Cov}\left(x(t_k), \tilde{y}(t_k^-)\middle|Y_0^{k-1}\right) \\ &\quad = \mathrm{E}\left((x(t_k) - \hat{x}(t_k^-))\tilde{y}(t_k^-)^{\mathsf{T}}\middle|Y_0^{k-1}\right) \\ &\quad = \mathrm{E}\left((x(t_k) - \hat{x}(t_k^-))\tilde{y}(t_k^-)^{\mathsf{T}}\right) \\ &\quad = \mathrm{Cov}(\tilde{x}(t_k^-), \tilde{y}(t_k^-)) = \mathrm{Cov}(\tilde{y}(t_k^-), \tilde{x}(t_k^-))^{\mathsf{T}} \,.\end{aligned} \tag{18e}$$

Using (17) with x, y, z replaced by $x(t_k), \tilde{y}(t_k^-), Y_0^{k-1}$ and applying (7) then gives

$$\begin{aligned}\mathrm{E}\left(x(t_k)\middle|\tilde{y}(t_k^-), Y_0^{k-1}\right) &= \hat{x}(t_k^-) \\ &\quad + \mathrm{Cov}\left(x(t_k), \tilde{y}(t_k^-)\middle|Y_0^{k-1}\right) \\ &\qquad \mathrm{Cov}\left(\tilde{y}(t_k^-)\middle|Y_0^{k-1}\right)^{-1}(y(t_k) - \hat{y}(t_k^-)) \,,\end{aligned} \tag{19a}$$

$$\begin{aligned}P(t_k^+) &= P(t_k^-) - \mathrm{Cov}\left(x(t_k), \tilde{y}(t_k^-)\middle|Y_0^{k-1}\right) \\ &\quad \mathrm{Cov}\left(\tilde{y}(t_k^-)\middle|Y_0^{k-1}\right)^{-1} \mathrm{Cov}\left(\tilde{y}(t_k^-), x(t_k)\middle|Y_0^{k-1}\right) \,,\end{aligned} \tag{19b}$$

which, by (18) and (7), results in the following simpler formulas:

$$\begin{aligned}\hat{x}(t_k^+) &= \hat{x}(t_k^-) \\ &\quad + \mathrm{Cov}(\tilde{x}(t_k^-), \tilde{y}(t_k^-))\,\mathrm{Cov}(\tilde{y}(t_k^-))^{-1}\tilde{y}(t_k^-) \,,\end{aligned} \tag{20a}$$

$$\begin{aligned}P(t_k^+) &= P(t_k^-) - \mathrm{Cov}(\tilde{x}(t_k^-), \tilde{y}(t_k^-)) \\ &\quad \mathrm{Cov}(\tilde{y}(t_k^-))^{-1}\,\mathrm{Cov}(\tilde{y}(t_k^-), \tilde{x}(t_k^-)) \,.\end{aligned} \tag{20b}$$

This completes the derivation of the basic formulas needed to explain the specific context in which the UT appears.

Remark 1: Equation (20) represents the measurement update (5), but in a more general formulation. Indeed, the standard formulation (5) of the linear CD-KF can be derived from (20) as follows:

$$\begin{aligned}y(t_k) &= Hx(t_k) + Du(t_k) + v(t_k) \\ \hat{y}(t_k^-) &= H\hat{x}(t_k^-) + Du(t_k)\end{aligned} \tag{21a}$$

giving $\tilde{y}(t_k^-) = H\tilde{x}(t_k^-) + v(t_k)$ and thus

$$\mathrm{Cov}(\tilde{x}(t_k^-), \tilde{y}(t_k^-)) = P(t_k^-)H(t_k)^{\mathsf{T}} \,, \tag{21b}$$

$$\mathrm{Cov}(\tilde{y}(t_k^-)) = H(t_k)P(t_k^-)H(t_k)^{\mathsf{T}} + R(t_k) \,. \tag{21c}$$

Hence, by defining

$$K(t_k) \triangleq \mathrm{Cov}(\tilde{x}(t_k^-), \tilde{y}(t_k^-))\,\mathrm{Cov}(\tilde{y}(t_k^-))^{-1} \,, \tag{22}$$

equation (5a) can be obtained by using (21b)-(21c) in (22), and equation (5b) can be obtained by using (22) in (20a). To derive equation (5c), first note that (18b) may be used to write (7c)–(7d) equivalently as

$$P(t_k^-) = \mathrm{Cov}\left(\tilde{x}(t_k^-)\middle|Y_0^{k-1}\right) \,, \tag{23}$$

$$P(t_k^+) = \mathrm{Cov}\left(\tilde{x}(t_k^+)\middle|Y_0^{k}\right) \,. \tag{24}$$

Now, use (20a) with (22) to obtain

$$\begin{aligned}\tilde{x}(t_k^+) &= \tilde{x}(t_k^-) - K(t_k)\tilde{y}(t_k^-) \\ &= (I - K(t_k)H(t_k))\tilde{x}(t_k^-) - K(t_k)v(t_k) \,,\end{aligned} \tag{25}$$

Then, applying (25) to (24), expanding and using (23) results in (5c).

The measurement update for the UKF corresponding to (5b)–(5c) in the linear case is given by (20a)–(20b), where the UT is used to estimate all the covariances and expected values. This is done by defining an ancillary function h_a by (26) for the measurement equation such that the stochastic part $\mathbf{x} \triangleq [x, v]$ and the deterministic part u are separated

$$\begin{aligned}y(t_k) &= h(x(t_k), u(t_k)) + v(t_k) \\ &\triangleq h_a([x(t_k), v(t_k)], u(t_k)) \,.\end{aligned} \tag{26}$$

The function $h_a(\cdot, u)$ plays the role of f in (12). Moreover, the definition of $\mathbf{x} = [x, v]$ and h_a does not in any way indicate a change in the model (8) but is only a way to formally explain exactly how the UT is used here. Given the function h_a and the input statistics

$$\mu_{\mathbf{x}} = \begin{pmatrix}\hat{x}(t_k^-) \\ \underline{0}\end{pmatrix} \,, \tag{27a}$$

$$C_{\mathbf{x}} = \begin{pmatrix}P(t_k^-) & \underline{\underline{0}} \\ \underline{\underline{0}} & R(t_k)\end{pmatrix} \,, \tag{27b}$$

the UT results in, via (15), the statistics

$$\mathrm{E}\left(y(t_k)\middle|Y_0^{k-1}\right) = \hat{y}(t_k^-) \,, \tag{28a}$$

$$\begin{aligned}&\mathrm{Cov}\left(y(t_k), x(t_k), v(t_k)\middle|Y_0^{k-1}\right) = \\ &\left[\mathrm{Cov}\left(y(t_k), x(t_k)\middle|Y_0^{k-1}\right) \quad \mathrm{Cov}\left(y(t_k), v(t_k)\middle|Y_0^{k-1}\right)\right] \\ &= \left[\mathrm{Cov}(\tilde{x}(t_k^-), \tilde{y}(t_k^-))^{\mathsf{T}} \quad \mathrm{Cov}\left(y(t_k), v(t_k)\middle|Y_0^{k-1}\right)\right] \,,\end{aligned} \tag{28b}$$

$$\mathrm{Cov}\left(y(t_k)\middle|Y_0^{k-1}\right) = \mathrm{Cov}(\tilde{y}(t_k^-)) \,, \tag{28c}$$

that are necessary in the UKF measurement update (20). Note that the additive measurement noise in (26) means that the UT only needs to include the stochastic effect related to x, while the effect of v can be added in a second step. The joint distribution of x, v in (27) is included here because it is necessary for the more general output equation $y(t_k) = h(x(t_k), u(t_k), v(t_k))$. Note also that the distribution used in (27) is conditional on Y_0^{k-1}, and that the omission of conditioning on Y_0^{k-1} in (28) is only correct in the linear CD-KF Gaussian case where the prediction errors $\tilde{y}(t_k^-)$ and the measurement Y_0^{k-1} are both uncorrelated and independent. However, this is an acceptable approximation in cases where the deviation from linearity and the Gaussian distribution is not severe.

When the UT is used for the time update, it must be formulated as a function of deterministic and stochastic inputs. The function

must represent the integration of the state SDE (29) from t_k to t_{k+1} starting from $\hat{x}(t_k^+), P(t_k^+)$, given $u(t), Q(t), t \in [t_k, t_{k+1}]$.

$$dx(t) = f(x(t), u(t))dt + dw(t) \ . \tag{29}$$

The SDE (29) can be simulated by discretizing in time. In general, the specific integration algorithm should relate to the interpretation of the SDE in the sense of either Ito or Stratonovich. However, since the diffusion (equal to 1 here) is independent of x, these two formalisms are identical [17, p.171]. The Euler-Maruyama integration [17], [21] may therefore be used as in (30).

$$\delta t_i \triangleq t_{i+1} - t_i \ , \ \delta w_i \triangleq w(t_{i+1}) - w(t_i) \ , \tag{30a}$$

$$\begin{aligned} &x(t_{i+1}) - x(t_i) \\ &\quad = f(x(t_i), u(t_i))\delta t_i + \delta w_i \\ &\quad = \left(f(x(t_i), u(t_i)) + \frac{\delta w_i}{\delta t_i} \right) \delta t_i \\ &\quad = \left(f(x(t_i), u(t_i)) + n_i \right) \delta t_i \\ &\qquad , \ n_i \in \mathrm{NID}(\underline{0}, Q(t_i)\delta t_i^{-1}) \ , \end{aligned} \tag{30b}$$

where i numbers the intermediate steps between t_k and t_{k+1}, and NID stands for Normally and Independently Distributed, i.e., white Gaussian noise. Similar to h_a for the measurement update, an ancillary function f_a is defined to explain the use of the UT in the time update. Using (30), the final state $x(t_{k+1})$ is a function

$$x(t_{k+1}) = f_a([x(t_k), N_{t_k}^{t_{k+1}}], U_{t_k}^{t_{k+1}}) \ , \tag{31}$$

of the initial state $x(t_k)$, stacked noise $N_{t_k}^{t_{k+1}}$ and input $U_{t_k}^{t_{k+1}}$. Note that $N_{t_k}^{t_{k+1}}$ and $U_{t_k}^{t_{k+1}}$ are the stacked noise n_i and input $u(t_i)$ for the intermediate time steps in the integration (30b), respectively. Additionally, note that x in (30)–(31) is an approximation of x in the SDE (29). Now, the UT transform can be used with the function f_a and the input statistics

$$\mathrm{E} \begin{pmatrix} x(t_k) \\ N_{t_k}^{t_{k+1}} \end{pmatrix} = \begin{pmatrix} \hat{x}(t_k^-) \\ \underline{0} \\ \vdots \\ \underline{0} \end{pmatrix} \ , \tag{32a}$$

$$\mathrm{Cov} \begin{pmatrix} x(t_k) \\ N_{t_k}^{t_{k+1}} \end{pmatrix} = \begin{pmatrix} P(t_k^-) & \underline{\underline{0}} & \dots & \underline{\underline{0}} \\ \underline{\underline{0}} & Q(t_k)/\delta t & \dots & \underline{\underline{0}} \\ \vdots & \vdots & \ddots & \vdots \\ \underline{\underline{0}} & \underline{\underline{0}} & \dots & Q(t_{k+1})/\delta t \end{pmatrix} \ , \tag{32b}$$

where the time step δt is assumed constant for simplicity. The result of the UT is then the output-input statistics

$$\hat{x}(t_{k+1}^-) \ , \ P(t_{k+1}^-) = \mathrm{Cov}(\tilde{x}(t_{k+1}^-)) \ , \tag{33}$$

that are necessary in the UKF time update corresponding to (6).

The choice of sub-sampling time δt is important; it must be chosen to be sufficiently small to guarantee that the errors from discretization are smaller than those from estimation. However, there is a tradeoff, as the number of sigma points from (32) given by $2(t_{k+1} - t_k/\delta t + 1) + 1$ increases with decreasing δt. For stiff systems in particular, a variable step-size integration is advantageous [22]. The presented method can easily change the sub-sampling step size δt at sampling times t_k; however, changing it between sub-samplings appears to be difficult. A detailed study of the effect of sub-sampling time is outside the scope of this paper; instead, the reader is referred to [22].

In summary, this section shows how to use the UT to calculate all the conditional means, variances and covariances necessary for the KF. This is why the method of this paper is called CD-UKF.

VI. Test of the CD-UKF

In this section, the CD-UKF developed in section V is tested by comparing it to two other methods using data from a simulated (nonlinear) Lorenz system. The first method is that described in [9], and the other method is the CD-EKF method presented in this paper.

From the theory, it is known that the methods will work perfectly for linear models. In particular, the output prediction errors (residuals) must be white noise. To rule out algorithmic errors, the methods developed by the authors of this paper have been successfully tested for linear models (the test results are not included here).

The comparison with the CD-UKF method of [9] and the CD-EKF is performed on the Lorenz system (34) that is also used by others for assessing filters [10]. Moreover, the deterministic part uses the same parameters (34f) as the original work by Lorenz [23].

$$dx = f(x)dt + dw \ , \tag{34a}$$

$$y = h(x) + v \ , \tag{34b}$$

$$f(x) = \begin{pmatrix} \sigma(x_2 - x_1) \\ x_1(\rho - x_3) - x_2 \\ x_1 x_2 - \beta x_3 \end{pmatrix} \ , \tag{34c}$$

$$h(x) = \begin{pmatrix} x_1 \\ x_3 \end{pmatrix} \ , \tag{34d}$$

$$w \in \mathrm{W}(0, Q) \ , \ v \in \mathrm{NID}(0, R) \ , \tag{34e}$$

$$\sigma = 10 \ , \ \rho = 28 \ , \ \beta = 8/3 \ , \tag{34f}$$

$$Q = 4.5^2 I \ , \ R = I \ , \tag{34g}$$

$$T_s = 0.01 \ , \ x(0) = \begin{pmatrix} 1 & 1 & 1 \end{pmatrix}^\mathsf{T} \ . \tag{34h}$$

The simulation runs from time 0 to 7 with a simulation time step of 0.0001 and an output sampling time of 0.01, resulting in 701 samples. Initially, 50 subsamples for the time update step are used for the CD-UKF method in [9]; however, this number could be reduced to 10 before the estimation RMS error started to increase. The sub sampling used for the developed algorithm is then adjusted to obtain the same performance, resulting in 2 subsamples. This choice is based on many simulation runs; the results of one run are shown in Figure 1–3, while Table III contains the statistics for 100 simulations.

Figure 1 shows the simulated states in 3D in blue; the measurements y_1, y_2 of the states x_1, x_3 are shown in red in the x_1, x_3 plane. The estimates of the developed CD-UKF method are shown in green. The output prediction errors are observed to be perfectly white in Figure 2. This is not guaranteed theoretically, but it does show that the filter works nearly optimally at least regarding the "linear" correlation. The so-called *lag dependence function* (LDF) [24], [25] estimates auto correlations from data, while also accounting for nonlinear dependence. As this is appropriate for the nonlinear system, it is also calculated, but not shown. The result is similar to those in Figure 2 except the non-zero lag correlations are slightly larger, though still small. The state estimation errors are shown in Figure 3 to illustrate that the uncertainty of the states x_1, x_3 is smaller and relatively stable compared to x_2, as expected, as there is no measurement of x_2 (the following shorthand notation $\tilde{x}_i^{\pm} \triangleq \tilde{x}_i(t^{\pm})$ and $\tilde{y}_i^- \triangleq \tilde{y}_i(t^-)$ is used in the figures and tables below).

Normal probability plots for the distribution of output and state prediction errors are shown in Figure 4–5. In the linear case, both will be Gaussian. For the Lorenz system, the output prediction errors appear to be nearly Gaussian, even though there are some deviations in the tails of the distribution. Such tail deviations are more pronounced for the state prediction errors, which then seem less Gaussian. This applies in particular to $\tilde{x}_2$, which has no "direct" measurement.

Table III shows the primary comparison results with two decimals of precision (the normalized values in all tables are obtained using

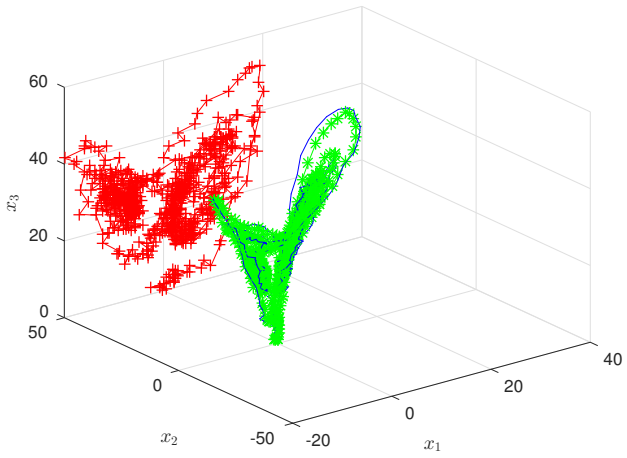

Fig. 1. Simulated states (blue) and measurements (red) for the Lorenz system. The estimated states using the developed CD-UKF are also included and are shown in green.

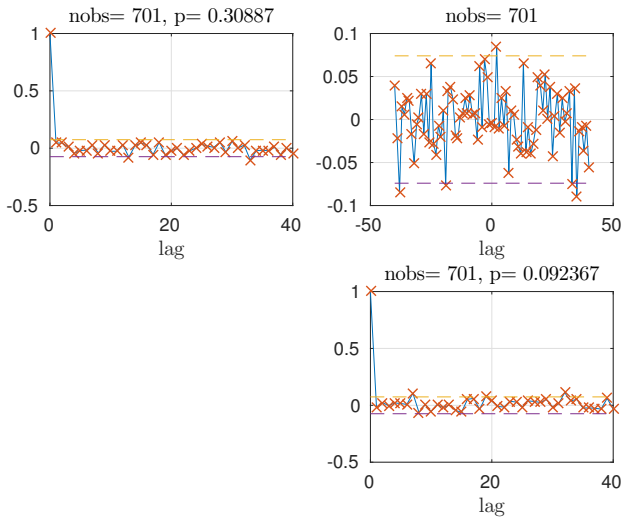

Fig. 2. Whiteness test for residuals from the CD-UKF. Upper left and lower right subplots show autocorrelations for $\tilde{\boldsymbol{y}}_1$ and $\tilde{\boldsymbol{y}}_2$, respectively (the $\boldsymbol{p}$ values are from the Portmanteau [26] whiteness test). The upper right plot shows the cross correlation between the two.

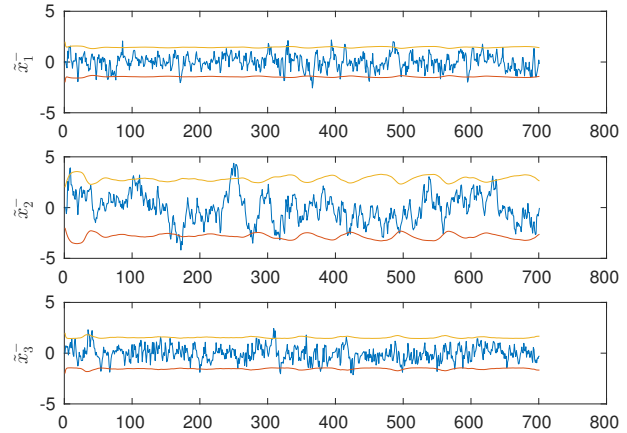

Fig. 3. State prediction errors (residuals) including 95% confidence limits from the CD-UKF, vs. time

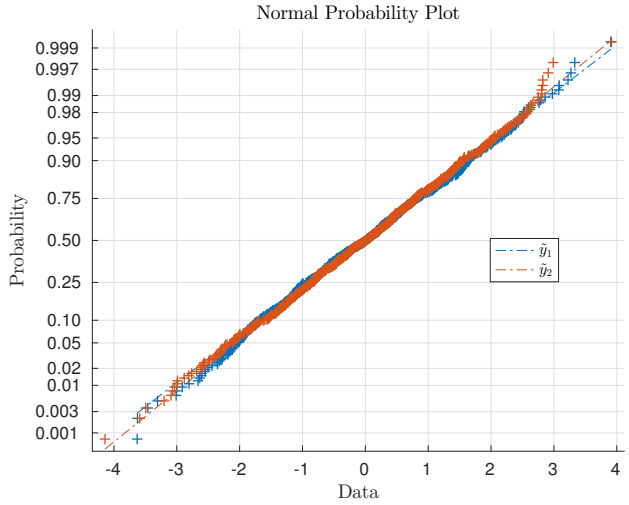

Fig. 4. Normal probability plots for output prediction errors (residuals) from the CD-UKF.

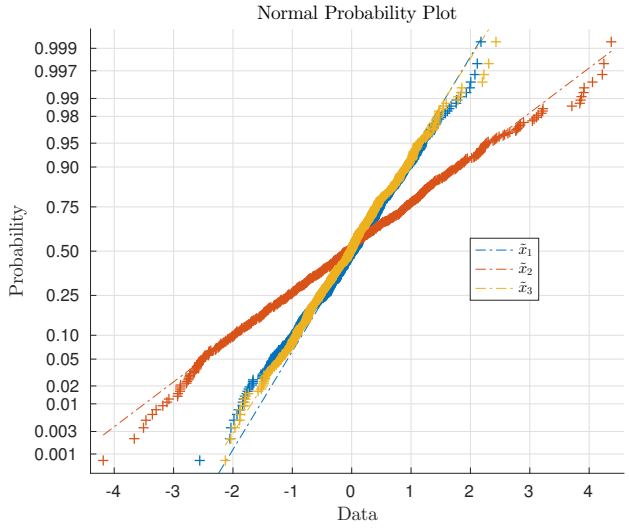

Fig. 5. Normal probability plots for state prediction errors (residuals) from the CD-UKF.

precise values for the numerator and denominator). It provides the mean and standard deviation based on 100 simulations of RMS errors, and the computation time (see also the Appendix.) As explained above, the accuracy values of the filters are made similar by tuning the number of subsampling time steps. The RMS of the state estimates using only the measurement would be 1, ∞ and 1 for x_1, x_2, x_3, respectively, due to the measurement (34d). This agrees well with the results in Table III, as the RMS values are below 1, ∞ and 1, respectively. The only difference in the methods' performances is that CD-UKF and CD-EKF are at least 9 times faster than the version reported in [9], denoted here by CD-UKFSS.

In many applications, the accuracy of the variance $\mathrm{Cov}(\tilde{x}_i(t_k^+))$ of the state estimate error $\tilde{x}_i(t_k^+)$ is as important as the state estimate itself. Therefore, this accuracy is also accessed by calculating the RMS of the normalized error

$$\tilde{x}_{n,i}(t_k^+) = \frac{\tilde{x}_i(t_k^+)}{\sqrt{P(t_k^+)_{ii}}} \tag{35}$$

with the estimated standard deviation $\sqrt{P(t_k^+)_{ii}}$ being the square root of the i^{th} diagonal element of $P(t_k^+) = \mathrm{Cov}\left(\tilde{x}_i(t_k^+)|Y_0^k\right)$. If

Mean						
Method	$\tilde{x}_1^+$	$\tilde{x}_2^+$	$\tilde{x}_3^+$	$\tilde{y}_1^-$	$\tilde{y}_1^-$	Time
CD-UKFSS	0.59	1.39	0.61	1.24	1.27	2.89
CD-EKF	0.58	1.37	0.61	1.23	1.27	0.37
CD-UKF	0.58	1.37	0.61	1.23	1.27	0.31
Normalized						
CD-EKF/CD-UKFSS	0.99	0.98	0.99	1	1	0.13
CD-UKF/CD-UKFSS	0.99	0.98	0.99	1	1	0.11
CD-UKF/CD-EKF	1	1	1	1	1	0.84
Standard Deviation						
CD-UKFSS	0.03	0.12	0.03	0.03	0.03	0.02
CD-EKF	0.03	0.11	0.03	0.03	0.03	0.01
CD-UKF	0.03	0.11	0.03	0.03	0.03	0.01
Normalized						
CD-EKF/CD-UKFSS	1	0.97	0.95	1	0.98	0.26
CD-UKF/CD-UKFSS	1	0.97	0.95	0.99	0.98	0.33
CD-UKF/CD-EKF	1	1	1	1	1	1.3

TABLE III

COMPARING THE PERFORMANCE OF ESTIMATION METHODS FOR THE LORENZ SYSTEM. THE MEANS AND STANDARD DEVIATIONS ARE BASED ON 100 SIMULATIONS. THE FIRST FIVE COLUMNS (IN BLOCKS ONE AND THREE) SHOW RMS OF STATE ESTIMATION ERRORS AND OUTPUT PREDICTION ERRORS, AND THE SIXTH COLUMN SHOWS THE COMPUTATION TIME.

the estimated standard deviation matches the actual standard deviation of the state estimation error, the normalized error $\tilde{x}_{n,i}(t_k)$ will have a standard deviation of one. Based on the 100 simulations used above, the estimated means and standard deviations of the RMS of $\tilde{x}_{n,i}(t_k)$ are shown in Table IV.

Mean			
Method	$\tilde{x}_{n,1}^+$	$\tilde{x}_{n,2}^+$	$\tilde{x}_{n,3}^+$
CD-UKFSS	1.24	1.06	0.89
CD-EKF	0.99	1	1
CD-UKF	0.99	1	1
Standard Deviation			
CD-UKFSS	0.06	0.08	0.04
CD-EKF	0.04	0.08	0.04
CD-UKF	0.04	0.08	0.04

TABLE IV

MEANS AND STANDARD DEVIATIONS OF RMS VALUES FOR NORMALIZED STATE ESTIMATION ERRORS. THE IDEAL VALUE OF THE MEAN IS ONE.

The values are close to one, especially for CD-EKF and CD-UKF, while the method of [9] shows a larger deviation, as the error variance is underestimated for the first component and overestimated for the third component.

As mentioned above, the values in the tables of this section are based on 100 simulations. Moreover, many other random simulation re-runs have been performed. The observed tendencies hold in general for the system (34) used for the test. It should, however, be noted that none of the algorithms have been optimized for speed.

VII. CONCLUSION

It is well known that the unscented Kalman filter performs at least as well as the extended Kalman filter for systems modeled by discrete time dynamic models and using discrete time measurements. A further advantage is that no linearization of state dynamics or measurement equations is needed. In many applications, it is preferable to use a dynamic model formulated in continuous time, while still having discrete time measurements. For this problem, [9] initially developed a method based on the unscented idea, where solving numerically the so-called Moment Differential Equations (MDEs) is necessary. The numerical solution of the MDEs has been extensively studied by others [12], [15], [16]. Compared to [9], this paper develops a more straightforward unscented method for the continuous-discrete estimation problem. Based on simulations, the proposed method is shown to result in the same RMS of state estimations errors as the present one while providing better covariance estimates; additionally, it runs approximately 9 times as fast.

APPENDIX

Minimum						
Method	$\tilde{x}_1^+$	$\tilde{x}_2^+$	$\tilde{x}_3^+$	$\tilde{y}_1^-$	$\tilde{y}_1^-$	Time
CD-UKFSS	0.53	1.08	0.56	1.17	1.18	2.86
CD-EKF	0.53	1.06	0.56	1.16	1.16	0.36
CD-UKF	0.53	1.06	0.56	1.16	1.16	0.3
Normalized						
CD-EKF/CD-UKFSS	1	0.98	1	1	0.98	0.13
CD-UKF/CD-UKFSS	1	0.98	1	1	0.98	0.11
CD-UKF/CD-EKF	1	1	1	1	1	0.84
Maximum						
Method	$\tilde{x}_1^+$	$\tilde{x}_2^+$	$\tilde{x}_3^+$	$\tilde{y}_1^-$	$\tilde{y}_1^-$	Time
CD-UKFSS	0.65	1.67	0.67	1.32	1.35	2.98
CD-EKF	0.64	1.64	0.67	1.32	1.34	0.39
CD-UKF	0.64	1.63	0.67	1.32	1.34	0.34
Normalized						
CD-EKF/CD-UKFSS	0.99	0.98	0.99	0.99	1	0.13
CD-UKF/CD-UKFSS	0.99	0.98	0.99	0.99	1	0.12
CD-UKF/CD-EKF	1	1	1	1	1	0.89

TABLE V

COMPARING THE PERFORMANCE OF ESTIMATION METHODS FOR THE LORENZ SYSTEM. THE MINIMUM AND MAXIMUM ARE BASED ON 100 SIMULATIONS. THE FIRST FIVE COLUMNS (IN BLOCKS ONE AND THREE) SHOW RMS OF STATE ESTIMATION ERRORS AND OUTPUT PREDICTION ERRORS, WHILE THE SIXTH COLUMN SHOWS THE COMPUTATION TIME.

Minimum			
Method	$\tilde{x}_{n,1}^+$	$\tilde{x}_{n,2}^+$	$\tilde{x}_{n,3}^+$
CD-UKFSS	1.12	0.84	0.78
CD-EKF	0.9	0.8	0.91
CD-UKF	0.9	0.8	0.91
Maximum			
CD-UKFSS	1.38	1.26	0.99
CD-EKF	1.1	1.17	1.1
CD-UKF	1.1	1.17	1.1

TABLE VI

RMS VALUES OF NORMALIZED STATE ESTIMATION ERRORS. THE MINIMUM AND MAXIMUM ARE BASED ON 100 SIMULATIONS.

ACKNOWLEDGMENT

The authors thank Simo Särkkä and Tronarp Filip from Department of Electrical Engineering and Automation (EEA), Aalto University, Finland for comments.

This work has partly been performed within the CL-Windcon project. This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 727477.

REFERENCES

[1] M. S. Grewal and A. P. Andrews, *Kalman Filtering*, 3rd ed. Wiley, Hoboken, New Jersey, 2008.

[2] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Journal of Basic Engineering*, vol. 82, pp. 35–45, 1960.

[3] A. H. Jazwinski, *Stochastic Processes and Filtering Theory*, ser. Mathematics in science and engineering, R. Bellman, Ed. Academic Press, Inc., 1970, vol. 64.

[4] S. J. Julier, J. K. Uhlmann, and H. F. Durrant-Whyte, "A new approach for filtering nonlinear systems," in *Proceedings of the IEEE 1995 American Control Conference*, vol. 3, Jun 1995, pp. 1628–1632.

[5] J. Uhlmann, "Dynamic map building and localization: New theoretical foundasions," Ph.D. dissertation, University of Oxford, 1995.

[6] S. Thrun, W. Burgard, and D. Fox, *Probabilistic Robotics*. The MIT Press, 2006.

[7] S. J. Julier and J. K. Uhlmann, "Unscented filtering and nonlinear estimation," *Proceedings of the IEEE*, vol. 92, no. 3, pp. 401–422, Mar 2004.

[8] E. A. Wan and R. V. D. Merwe, "The unscented kalman filter for nonlinear estimation," in *Proceedings of the IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium (Cat. No.00EX373)*, 2000, pp. 153–158.

[9] S. Särkkä, "On unscented kalman filtering for state estimation of continuous-time nonlinear systems," *IEEE TRANSACTIONS ON AUTOMATIC CONTROL*, vol. 52, no. 9, pp. 1631–1641, 2007.

[10] G. Evensen, "Advanced data assimilation for strongly nonlinear dynamics," *American Meteorological Society*, vol. 125, pp. 1342–1354, 1997.

[11] ——, "The ensemble kalman filter: theoretical formulation and practical implementation," *Ocean Dynamics*, vol. 53, no. 4, pp. 343–367, 2003.

[12] G. Kulikov and M. Kulikova, "Accurate continuous-discrete unscented kalman filtering for estimation of nonlinear continuous-time stochastic models in radar tracking," *Signal Processing*, vol. 139, pp. 25–35, October 2017, available online 4 April 2017.

[13] M. V. Kulikova and G. Y. Kulikov, "On computational robustness of accurate continuous-discrete unscented kalman filtering for target tracking models," in *Proceedings of the 2016 European Control Conference*, 2016, pp. 1129–1134.

[14] M. Takeno and T. Katayama, "A numerical method for continuous-discrete unscented kalman filter," *Special Issue of International Journal of Innovative Computing, Information and Control*, vol. 8, no. 3, pp. 2261–2274, 2012.

[15] S. Kadu, M. Bhushan, and K. Roy, "Continuous discrete unscented kalman filtering for nonlinear differential algebraic equations systems," *Computer Aided Chemical Engineering*, vol. 31, pp. 940–944, 2012.

[16] S. Lv, J. Chen, and Z. Liu, "UDU^T continuous-discrete unscented kalman filtering," in *Second International Symposium on Intelligent Information Technology Application*, 2008, pp. 876–879.

[17] L. Arnold, *Stochastic differential equations: theory and applications*. Wiley-Interscience [John Wiley & Sons], New York-London-Sydney, 1974, translated from the German.

[18] X. Luo and I. Moroz, "Ensemble kalman filter with the unscented transform," *Physica D*, vol. 238, pp. 549–562, 2009.

[19] R. van der Menve and E. A. Wan, "The square-root unscented kalman filter for state and parameter-estimation," in *2001 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No.01CH37221)*, vol. 6, IEEE. IEEE, 2001, pp. 3461–3464.

[20] S. Julier, "A skewed approach to filtering," in *The Proceedings of the 12th Intl. Symp. On Aerospace/Defense Sensing, Simulation and Controls*, vol. 3373. SPIE–The International Society for Optical Engineering, 1998, pp. 1–12.

[21] P. E. Kloeden and E. Platen, *Numerical Solution of Stochastic Differential Equations*, ser. Stochastic Modelling and Applied Probability. Springer-Verlag Berlin Heidelberg, 1992, vol. 23.

[22] G. Y. Kulikov and M. V. Kulikova, "Estimating the state in stiff continuous-time stochastic systems within extended kalman filtering," *SIAM Journal on Scientific Computing*, vol. 38, no. 6, pp. A3565–A3588, 2016.

[23] E. N. Lorenz, "Deterministic nonperiodic flow," *Journal of the Atmospheric Sciences*, vol. 20, no. 2, pp. 130–141, 1963.

[24] H. A. Nielsen and H. Madsen, "A generalization of some classical time series tools," *Computational Statistics & Data Analysis*, vol. 37, pp. 13–31, 2001.

[25] E. Lindström, "Semiparametric lag dependent functions," *Applied Mathematical Sciences*, vol. 7, no. 12, pp. 551–566, 2013.

[26] H. Madsen, *Time Series Analysis*. Chapman & Hall, 2008, iSBN: 978-1-4200-5967-0.